DOCUMENT RESUME

ED 218 798 EA 014 828

AUTHOR Gottschalk, Rand; And Others
TITLE Multi-Site Implementation and Replication of Research Findings: Is the Modified RD&D Model Viable?
SPONS AGENCY National Science Foundation, Washington, D.C.
PUB DATE 1 Oct 81
GRANT ISI-7920576-01
NOTE 21p.; Paper presented at a meeting of the Evaluation Research Society (Austin, TX, October 1, 1981).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Adoption (Ideas); Change; Change Strategies; Elementary Secondary Education; Innovation; Program Effectiveness; *Program Implementation
IDENTIFIERS Program Modification; *Program Replication; *Research Development and Dissemination

ABSTRACT

Described here is a research project being undertaken to address two points: the relationship of the fidelity of program replication to program outcomes at user sites; and the extent to which fidelity can occur in programs with more than a few replicates. The findings will have bearing on the usefulness of the Research, Development, and Diffusion (RD & D) Model, which has been criticized on the grounds that it does not take into account the modification or "reinvention" of programs that occurs when they are replicated at other sites. In the current study, eight programs are being examined--four in education and four in the criminal justice field. Fidelity of replication is to be determined using Hall and Loucks's methods of measuring the number and importance of program components implemented. Data have already been collected by telephone interview with program developers and implementers and will also be gathered during site visitations. The relationship between fidelity and program outcomes will be analyzed using a multiple regression approach. Although the research is at an early stage of data collection, it can already be concluded that specificity in detailing program components is critical in conducting both evaluation and implementation research. (Author/JM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED218798

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Craig Blakely

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

MULTI-SITE IMPLEMENTATION AND REPLICATION OF RESEARCH FINDINGS: IS THE MODIFIED RD&D MODEL VIABLE?

Rand Gottschalk
David Roitman
James Emshoff
Craig Blakely
Michigan State University

Paper presented at the Evaluation Research Society, Austin, Texas, October 1, 1981. This work was completed under a grant from the National Science Foundation # ISI-7920576-01.

EA 014 828

To what extent is the fidelity of a social program's implementation related to the outcomes that program produces? Can social technologies be disseminated on a large scale such that they are implemented with fidelity? These two questions are of increasing importance given the recent movement towards decentralization of program initiatives in the present administration.

This decentralization is largely based on political motives. However, a body of research exists to support the decentralization of social programming. This research points out the general failure of efforts based on the "Research, Development, and Diffusion model" or RD&D (Havelock, 1969) to succeed in fostering program implementation. The RD&D model, which guided much of the Federal social programming of the 1960's, has five main features (House, Kerins, & Steele, 1972):

1)There is a rational sequence of activities which moves from research to development to packaging.

2)Planning must occur on a large scale.

3)A division of labor separates roles and functions in the overall process.

4)A passive consumer awaits acceptance of the innovation if it is delivered properly.

5)A high initial development cost is necessary to eventual success.

Much recent research has found fault with this model. The futility of disseminating complex social innovations as "canned

packages" through brochures and other written materials has been repeatedly demonstrated(Fairweather, Sanders, & Tornatzky, 1974; Fullan & Pomfret, 1977). Treating practitioners as passive consumers has also been demonstrated to be a waste of time and money (Gross, Giacquinta, & Bernstein, 1971). Practitioner involvement in implementation decisions seems to be a virtual prerequisite for successful implementation (Fullan & Pomfret, 1977; Tornatzky, Fergus, Avellar, & Fairweather, 1981). Exacerbating these problems is the fact that programs often undergo some modification or "reinvention" by users in order to accomodate programs to specific sites (Rice & Rogers, 1980).

However, the momentum of the modification or "reinventionist" movement may be impelling us to throw the RD&D baby out with the bath water, a position cogently stated by Datta (1981). A good portion of Datta's arguement rests on her critique of the large scale RAND study of federal programs supporting educational change (Berman & McLaughlin, 1978). The RAND research is perhaps the most influential of the reinventionist studies. According to Datta, the RAND findings that "mutual adaptation"(of site and program) generally characterizes successful implementation is suspect. Although the authors claimed that actual program replication rarely occured, their implementation outcome measure was "the extent to which projects met their own goals" (Berman & McLaughlin, 1977, Vol. VII, p. 50). This definition of implementation builds the occurrence of adaptation into the results. Secondly, the "programs" examined

by the RAND study were never specified sufficiently to enable "high fidelity" implementation. Finally, the "massive infusion of federal dollars", reputed to characterize the Office of Education programs which were studied averaged $100.00 per pupil per year across programs. This is hardly a great expense when the extent and the complexity of the changes which were to be produced is considered.

In summary, Datta concluded that the RD&D model for social programming was not directly tested by the RAND research despite claims to the contrary. Similar criticisms can be levelled at other research supporting the modification point of view. Measures of implementation are frequently too flawed to substantiate research claims (Scheirer, 1981), and program parameters are often left poorly specified or entirely unspecified (Eveland, Rogers, & Klepper, 1977).

It is true, that the milieu of social programs is hardly as rational a place as was previously assumed. Authors who have pointed out the need for disseminators and administrators to take interpersonal and organizational dynamics into account have made valuable contributions to the study of implementation (e.g. Yin, 1978). However, it may be that a need still exists to develop programs through careful and valid research, and to disseminate programs which have been demonstrated to attain beneficial outcomes for clients with sufficient fidelity to produce these same outcomes at the user sites.

We would argue that the connotations of "research, develop-

ment, and demonstration" need not include those assumptions demonstrated to be inadequate. Replacing policies based on these assumptions with policies supported by reinventionist research produces a modified RD&D approach. This approach encourages research and development by practitioners, utilizes active, full-time change agents to disseminate rigorously evaluated programs, and pays attention to organizational-environmental dynamics shown by past research to be crucial for successful implementation. This modified RD&D approach, presently utilized by two federal agencies, is the focus of the present research.

The intent of this research has been to develop highly-specific, behavioral descriptions of eight social programs disseminated by these agencies. These descriptions will then be used to address the two issues posed at the beginning of this paper: 1) the relationship of fidelity to program outcomes at user sites; and 2) the extent to which fidelity can exist in programs with more than a few replicates. By using these highly-specified behavioral descriptions, it is hoped that the findings of this research will bear more directly on questions concerning the usefulness of the RD&D model than previous studies in this area.

Program Fidelity

The development of the behavioral instruments followed a methodology pioneered by Hall, Loucks, and their associates(Heck, Stiegelbauer, Hall, & Loucks, 1981), working in the area of educational innovation. These researchers have conceptualized

social programs as consisting of a finite number of components or parts. Program fidelity may be defined following this schema in terms of the number of components actually implemented at a site. This may be expressed as a percentage (number of components implemented/number of components in the program). Comparisons can then be made both within program and with other programs with different numbers of components.

A second major concept in the Hall and Loucks methodology involves the importance of various components. Some components are more important(to various actors, for various purposes) than other components. Hall and Loucks have contrasted the core components of a program with its related components. (We prefer the terminology essential vs non-essential in order to tie the component to an exogenous referent such as outcomes). Also, components can be implemented in a variety of ways. Hall and Loucks call this phenemenon "component variation". These variations may be considered ideal, acceptable, or unacceptable by various actors (disseminators, administrators, practitioners) depending on their perspective. Variations of components are implemented in different identifiable patterns or configurations.

Hall and his associates have used this framework to study the implementation of numerous educational innovations. To our knowledge, the research discussed in this paper is the first to use this methodology in other areas of social programming outside of education. One other study has used this methodology to examine federal-level dissemination activity (Network, 1981).

However this study was also concerned solely with education. In addition, these researchers have chosen to use a different level of component specificity than that utilized in the present research. Our program descriptions range from 36-105 components, while those of the Network, and Hall and his associates are generally in the 5-20 range.

Research Methodology

Overview

Our research is a three-phased effort. The first phase involved the identification of adopters through a semi-structured phone interview concerning their reasons for program adoption. The second phase of the research, which is currently in progress, involves phone interviews with the adopters identified in the first phase. The second phase is concerned with measuring the degree of program implementation and routinization. The third phase will involve site visits to a sample of the organizations interviewed during the second phase.

Innovation Selection

Programs from two federal agencies, one agency dealing with education, and one agency in the criminal justice field, were selected. The educational agency was the National Diffusion Network(NDN). The goal of this agency is the widespread adoption of innovations that are validated by the Joint Dissemination Review Panel. The NDN uses both innovation developers and regional facilitators as change agents. The review process acts as a quality control on the innovations (Emrick, Peterson, & Agarwala-

Rogers, 1977). In criminal justice, the Law Enforcement Assistance Administration Exemplarary Projects Program also requires on-site evidence of effectiveness before a program is designated as exemplarary and information is disseminated. Both agencies only disseminate information on programs that are known to have produced desirable outcomes at the original site. Variation in effectiveness across replicates of a given program can then be attributed to differences in program implementation at the sites.

The first step in selecting innovations for study involved the reading of materials describing all the NDN programs(about 150 programs) and LEAA exemplarary projects(about 35 programs). Four innovations from each field were chosen that met the following criteria: innovations had to be in the dissemination phase for at least two years, innovations had to have a reasonable chance of having at least 20 replicates, and innovations had to be "organization-wide"(either involve multiple units of the organization and/or demand a good deal of interaction with the organization's environment). Information on replicates of the innovations were obtained from two sources. Developers of the innovation were contacted regarding information they had concerning adopters of their programs. The second source of information was a three percent random sample selection of the appropriate organization unit(police department, court, prison, or school). Adopters were located through a semi-structured interview. In this phase of the research, the "organization" was considered to be the implementing unit. In the second phase of our

research, the "organization" was considered to be the unit in which the innovation was housed. The organization can thus differ between Phase I and Phase II of our research in the criminal justice area. As J.D. Eveland(1981) has remarked, the choice of the appropriate organizational unit is very important. Our experience substantiates this point. For this research, consistency across innovations in defining the unit of analyses is required in order to enable across-innovation analyses. Maintaining consistency was therefore the first priority in the operational definition of "organization".

Development of the Implementation Measure

The second phase of our research involves an additional interview with each adopter previously identified in Phase I. This interview focused on the implementation of the innovation in the adopting organization. Other issues that we are investigating which will not be discussed in this paper involve the concept of routinization. Of special interest to us is the magnitude and type of relationship between implementation and routinization.

In order to measure program implementation it is first necessary to specify what the program is and what it involves. To do this, we are using what we hope to be an ideal (or at least an acceptable) variation of the methodology used by Hall and his associates. The first step entailed reading all written information on each of the programs. Two staff members then visited each innovation developer. These researchers observed either the

original site of the innovation or the site used for dissemination purposes. During these visits, intensive conversations were held with the developers. These conversations were recorded on tape. These conversations concerned what one would likely see if he or she visited an innovation adopter. This type of question, used by Hall and his associates, focuses the developer's attention on the observable behaviors the innovation entails. Discussion also included how the typical client of the innovation would interact with various other actors involved in the innovation. At most of these site visits, the programs were no longer being run directly by the original developers of the program, and conversations with administrators, front-line staff, and clients were also recorded. This procedure follows the concern of Hall and his associates that descriptions of programs for the purposes of implementation measurement should not be limited to the developer's perspective.

The second step in our process involved the separation of all the information we had obtained into distinct, identifiable components. This process involved one of the staff who had gone on the developer/site visit and one additional researcher. The first part of this process entailed re-reading all the program materials. The tapes of the conversations were heard by both the staff who had gone on the visit and an additional staff member. While listening to the tape or reading the materials, all identifiable activities were put onto cards. These cards were then sorted into categories of innovation-specific activities. Dup-

lication between activities were eliminated. One of the people who had visited the developer then took responsibility for the development of the rest of the measure. This involved eliminating further redundancies in the activities and making sure that each component entailed only one or a few related activities. It is at this point that our measurement development started to diverge from the strategies used by Hall and his associates. Due to the limited number of program adopters we were working with, it was decided to not interview users of the program to obtain the variations. Instead, one of the people who did the site visit generated the variations. These were reviewed by other researchers on the team. These initial component lists and variations ranged in length from 40-116 components. Extreme care was taken to insure a common level of specificity across innovations. Thus, variability in the actual number of components identified for each program represents true variability in the level of the complexity of the innovation.

These component lists were then sent to the developers. The developers were given two tasks. The first was to respond to a list of the components. Developers were asked to label the component as either "essential", "non-essential", or "not relevant" to their program. That is, did the developer view the component as critical or necessary for the developer to construe an adopter as a true replicate of the initial innovation? Developers were also asked to decide whether the component was required in order to obtain the validated outcomes for the program. The second

task that we asked developers to undertake was the rewriting of our variations, based on their experience and thoughts concerning what constituted an ideal, acceptable, or unacceptable variation. Data-gathering involved a phone interview with developers which covered every variation of every component. It should be mentioned that in cases where there was more than one developer at a site, each developer was asked to assist in these two tasks. Hall and his associates have argued cogently for the importance of taking multiple perspectives into account in this process. Therefore, for one-half of the innovations, program users at the original site were asked to perform the same tasks as the developers. In revamping the instrument to reflect respondent's comments, every effort was made to preserve both perspectives. In addition, components were "tagged" from both the developer and users' view, as to whether the component was essential, non-essential, or not relevant, and either important or unimportant to outcomes.

Thus, eight separate fidelity interviews were constructed with their own component and variation lists. Example component items from several innovations are included in Table 1. Readers should note the similar level of specificity across program categories. We are just completing the piloting of this phase of the research.

The interview strategy we have been using is fairly open-ended. We ask the user of the innovation about their use of a particular category of activity (a set of components). As they

talk, we code the variation they are using under the appropriate component. After we have coded every component in a category, we ask the user whether we have missed anything, and whether they are doing anything "different or unique" with regard to that set of components. This is then noted as the amount of modification that has taken place in that category.

Respondents in this phase are users who have been identified as most familliar with the day to day running of the program. In addition, inter-coder reliability is assessed by having two researchers code a sample of interviews. To obtain a measure of interrespondent validity, a second respondent will be interviewed at a sample of sites.

Site Visits and Collection of Outcome Data

A 50% sample of the adopters of each of the innovations will be site-visited following the telephone interviews. These adopters will be chosen from the high and low regions of the fidelity range obtained from these interviews. Fidelity will be calculated in a number of different ways. First, the percentage of components implemented/the total number of components will be computed. This score will enable comparisons across programs. A program-specific calculation of fidelity will also be carried out by assigning a score of 2 for components implemented ideally, 1 for components implemented acceptably, and 0 for components implemented unacceptably. Both these calculations will be used to select high and low fidelity sites to visit.

Site visits will not be exact replications of phone inter-

views, since changes in implementation status can occur in the interim between the phone interview and the site visit. We will try to minimize the impact of these changes on our measures by asking respondents what has happened on-site since we talked to them last. The site visit will then be used to contrast results of observational methods of data collection with those of phone interview methods.

A second major purpose of the site visits will be the collection of outcome data for each program. These will be the original validated outcomes of the program. Many of the innovations require their adopters to collect this outcome data.

Data Analysis and Questions to be Answered

The major questions to be answered concern the relationship between fidelity and outcomes, and whether certain configurations of components produce higher levels of outcomes than other configurations. The relationship between fidelity and outcomes will be analyzed using a multiple regression approach. Program effectiveness (outcome levels) becomes the criterion variable and each of the components become the predictors. The components could be considered in a binary sense (presence or absence) or as a three-point continuum (ideal=2, acceptable=1, unacceptable=0). The standardized beta weights for each of the components will indicate the importance of that particular component to the outcome(s). The multiple R will indicate the magnitude of the relationship between implementation and program effectiveness.

The innovation configurations cannot be scaled. However, a

non-parametric approach can be used. Each configuration can be given a different nominal code number and outcome scores can be divided into ordinal categories of high, medium, and low effectiveness. Chi-square analyses can be used to determine the relative frequencies of various configurations for different outcome levels.

It will be recalled that data were collected from developers concerning which components they considered important for producing outcomes. The above analysis will suggest which components are empirically important for producing outcomes. These two sources will then be compared in order to determine to what extent developers are aware of components essential to produce outcomes. If developers are not aware of which components produce outcomes, this would suggest major faults in even a modified RD&D model, since disseminators would presumably not emphasize "non-essential" components. The above chi-square analysis would speak to the usefulness of the concept of configurations.

Summary and Conclusions

The wave of recent research and commentary critical of the classical RD&D policy model has already influenced program policies to a considerable extent. Although many of the points raised in the "reinventionist" literature are well taken, the essential assumptions of the RD&D model have not been directly addressed by research. Our research is an early step in this direction.

Unfortunately this research is at an early stage of data collection and we cannot yet directly address the issues raised here today. However, we have already learned a great deal from the process. Some of what we have learned relates directly to J.D. Eveland's(1981) earlier comments concerning the communalities between evaluation research and implementation research. Specifically, specificity in detailing program components or innovation configurations is critical. The social scientist conducting implementation research in a post hoc manner similar to our project is extremely limited by the detail of the initial innovation's evaluation and the degree of specificity with which the dissemination efforts were undertaken. The relationship between implementation and outcome must be clearly detailed at each step along the way. The demands of this area are such that they render previous research paradigms somewhat impotent in areas of interest to the applied and/or implementation researcher. The black box that is a successful innovation is no longer acceptable in and of itself. A hybrid evaluation model that clearly ties specific innovation components and configurations to specific program outcomes is essential for the implementation researcher, innovation developer, program evaluator, program disseminator, and program adopter.

References

Berman, P. & McLaughlin, M.W. Federal programs supporting educational change, Vol. 7: Factors affecting implementation and continuation. April, 1977, The Rand Corporation, Grant No. R-1589/7-HEW, U.S. Office of Education.

Berman, P. & McLaughlin, M.W. Federal programs supporting educational change, Vol. 8: Implementation and sustaining innovations. Final Report, May 1978, The Rand Corporation, Grant No. r-1589/8-HEW, U.S. Office of Education.

Datta, L.E. Damn the experts and full speed ahead: An examination of the study of federal programs supporting educational change, as evidence against directed development and for local problem-solving. Evaluation Review, 1981, 5(1), 5-32.

Emrick, J.A., Peterson, S.M., & Agarwala-Rogers, R. Evaluation of the National Diffusion Network, Vol. 1: Findings and recommendations (SRI Project 4385). Washington, D.C.: U.S. Office of Education, Department of Health, Education, and Welfare, May 1977.

Eveland, J.D. Evaluating the implementation of organizational technology. Paper presented at the joint meeting of the Evaluation Network and Evaluation Research Society, October, 1981.

Eveland, J.D., Rogers, E., & Klepper, C. The innovation process in public organizations: Some elements of a preliminary model. Springfield, VA: NTIS, March, 1977.

Fairweather, G.W., Sanders, D.H., & Tornatzky, L.G. Creating change in mental health organizations. New York: Pergamon Press, 1974.

Fullan, M., & Pomfret, A. Research on curriculum and instruction implementation. Review of Educational Research, 1977, 47, 335-397.

Gross, N., Giaquinta, J., & Bernstein, M. Implementing educational innovations. New York: Basic Books, 1971.

Hall, G.E. & Loucks, S.F. Innovation Configurations: Analyzaing the adaptation of innovations. Austin: Research and Development Center Teacher Education, The University of Texas, 1978.

Havelock, R.G. Planning for innovation through dissemination and utilization of knowledge. Ann Arbor: Institute for Social Research, 1969.

Heck, S., Steigelbauer, S., Hall, G.E., & Loucks, S.F. Measuring innovation configurations: Procedures and applications. Austin: Research and Development Center for Teacher Education, The University of Texas, 1981.

House, E.R., Kerins, T., & Steele, J.M. A test of the research and development model of change. Educational Administration Quarterly, 1972, 8, 1-14.

Network, Inc. A study of dissemination efforts supporting school improvement. Andover, Mass: The Network, Inc., 1981.

Rice, R.E., & Rogers, E.M. Reinvention in the innovation process. Knowledge: Creation, Diffusion, Utilization, 1980, 1, 499-514.

Scheirer, M.A. Defining utilization of innovations: Issues in the measurement of implementation. Paper presented at the Evaluation Research Society, Austin, 1981.

Tornatzky, L.G., Fergus, E.O., Avellar, J.W., Fairweather, G.W., & Fleischer, M. Innovation and social process: A national experiment in implementing social technology. New York: Pergamon Press, 1981.

Yin, R.K., Quick, S.K., Bateman, P.M., & Marks, E.L. Changing urban bureaucracies: How new practices become routinized. Santa Monica, CA: The RAND Corporation, 1978.

TABLE 1

Example Innovation Components

Assessment of Clients Interests.

Criminal Justice.

Skills/Interests Assessed.

I. Residents skills/interests are assessed prior to beginning training and job search.

U. Residents skills/interests are not a major concern. Jobs are considered as they are available.

Education.

Career Abilities and Interests: Individualized, Systematic Assessment.

I. Student career abilities and interests are individually and systematically assessed using explicit procedures.

U. Student career abilities and interests are not assessed individually or systematically.

U. Student career abilities and interests are not assesed at all.

Regular Monitoring of Clients Progress.

Criminal Justice.

Monitoring of Residents' Progress.

I. Once a month residents are behaviorally rated concerning contracted progress.

A. Every other month residents are behaviorally rated concerning contracted progress.

U. Every third month or less, residents are behaviorally rated concerning contractual progress.

U. Residents are not regularly rated concerning their contractual progress.

Education.

Academic Learning Plan: Feedback/Progress Sessions.

I. Regular (periodic) 1-to-1 feedback and progress meetings are held with students with reference to the Academic/ Learning plan, at least once a week.

A. Regular (periodic) 1-to-1 feedback and progress meetings are held with students at least once a month.

U. Regular (periodic) 1-to-1 feedback and progress meetings with reference to Academic/Learning plan are held with students less than once a month.